CLAIMS

1. A computer-based method of drug design based on genetic polymorphisms, comprising:

obtaining more than one amino acid sequence of target proteins that are the product of a gene exhibiting genetic polymorphisms, wherein the sequences represent different genetic polymorphisms;

generating 3-dimensional (3-D) protein structural variant models from the sequences; and

based upon the structures of the 3-D models, designing drug candidates, modifying existing drugs, identifying potential drug candidates or identifying modifications of existing drugs based on predicted intermolecular interactions of the drug candidates or modified drugs with the structural variants.

2. The method of claim 1, wherein the structure-based drug design method comprises:

computationally docking the drug candidate or modified drug molecules with the target protein structural variant models;

energetically refining the docked complexes;

determining the binding interactions between the drug candidate or modified drug molecules and the structural variants; and

designing and identifying drugs or modifications to existing drugs based on the binding interactions.

3. The method of claim 2 wherein the binding interactions are determined by:

calculating the free energy of binding between the protein structural variant model and the docked molecule; and

decomposing the total free energy of binding based on the interacting residues in the protein active site.

4. The method of claim 1 wherein:

after the protein structural variant models derived from a particular genetic polymorphism are generated, selected model structures are

analyzed to determine common structural features that are conserved throughout the selected models, wherein

the conserved structural features are used as a basis for structure-based drug design studies.

5. The method of claim 4, wherein the conserved structural features are stretches of non-contiguous residues, wherein each stretch contains at least two amino acids.

6. The method of claim 5, wherein the protein is human immunodeficiency virus protease.

7. The method of claim 6, wherein the conserved residues comprise residues comprise residues 1-9, 25-29, 49-52, 78-81 and 94-99; and wherein:

residue 1 is an aliphatic amino acid; residue 2 is a hydrophilic amino acid; residue 3 is an aliphatic amino acid; residue 4 is a hydrophilic amino acid; residue 5 is a hydrophobic amino acid; residue 6 is an aromatic amino acid; residue 7 is a hydrophilic amino acid; residue 8 is a basic amino acid; residue 9 is an aliphatic amino acid; residue 25 is an acidic amino acid; residue 26 is a hydrophobic amino acid; residue 27 is an aliphatic amino acid; residue 28 is an aliphatic amino acid; residue 29 is an acidic amino acid; residue 49 is an aliphatic amino acid; residue 50 is a hydrophobic amino acid; residue 51 is an aliphatic amino acid; residue 52 is an aliphatic amino acid; residue 78 is an aliphatic amino acid; residue 79 is an aliphatic amino acid; residue 80 is a hydrophilic amino acid; residue 81 is an aliphatic amino acid; residue 94 is an aliphatic amino acid; residue 95 is a thio-containing amino acid; residue 96 is a hydrophilic amino acid; residue 97 is hydrophobic amino acid; residue 98 is hydrophilic amino acid; and residue 99 is an aromatic amino acid.

8. The method of claim 6, wherein the conserved residues comprise residues comprise residues 1-9, 25-29, 49-52, 78-81 and 94-99; and wherein:

residue 1 is proline; residue 2 is glutamine; residue 3 is isoleucine; residue 4 is threonine; residue 5 is leucine; residue 6 is tryptophan; residue 7 is glutamine; residue 8 is arginine; residue 9 is proline; residue 25 is aspartic acid; residue 26 is threonine; residue 27 is glycine; residue 28 is alanine; residue 29 is aspartic acid; residue 49 is glycine; residue 50 is isoleucine; residue 51 is glycine; residue 52 is glycine; residue 78 is glycine; residue 79 is proline; residue 80 is threonine; residue 81 is proline; residue 94 is glycine; residue 95 is cysteine; residue 96 is threonine; residue 97 is leucine; residue 98 is asparagine; and residue 99 is phenylalanine.

9. The method of claim 6, wherein the HIV protease has the sequence of amino acids set forth in any of SEQ ID Nos. 3-74 and 77-117.

10. The method of claim 9, wherein the residues comprise residues 1-9, 25-29, 49-52, 78-81 and 94-99.

11 ~~10~~. The method of claim 1, wherein the selected model structures represent the structural variants resulting from the most commonly occurring genetic polymorphisms.

12 ~~11~~. The method of claim 1, wherein the selected model structures represent the structural variants resulting from genetic polymorphisms found in a selected patient subpopulation.

13 ~~12~~. The method of claim 1 wherein the structural variant models are stored in a relational database, comprising:

3-D molecular coordinates for the structural variants;

a molecular graphics interface for 3-D molecular structure visualization; computer functionality for protein sequence and structural analyses; and

database searching tools.

14 ~~13~~. The method of claim ~~12~~ 13, wherein the database further comprises one or more of observed clinical data associated with the genetic polymorphisms, subject medical history and subject history.

15 ~~14~~. The method of claim 1, wherein:

after generating the 3-D protein structural variant models, the method comprises:

computationally docking drug molecules with the target protein models; and

energetically refining the docked complexes; and

wherein the candidate drugs are specific for a protein with a selected polymorphism or specifically interact with all proteins exhibiting a polymorphism.

16. The method of claim 15, wherein the structure-based drug design method comprises:

computationally docking drug or potential new drug candidate molecules with the target protein structural variant models;

energetically refining the docked complexes;

determining the binding interactions between the drug or potential new drug candidate molecules and the structural variants; and

designing potential new drugs or modifications to existing drugs based on the binding interactions.

17. The method of claim 16, wherein the binding interactions are determined by:

calculating the free energy of binding between the protein structural variant model and the docked molecule; and

decomposing the total free energy of binding based on the interacting residues in the protein active site.

18. The method of claim 15, wherein:

after the protein structural variant models derived from a particular genetic polymorphism are generated, selected model structures are analyzed to determine common structural features that are conserved throughout the selected models; and

the conserved structural features are used as a basis for structure-based drug design studies.

18. The method of claim 17, wherein the selected model structures represent the structural variants resulting from the most commonly occurring genetic polymorphisms.

19. The method of claim 17, wherein the selected model structures represent the structural variants resulting from genetic polymorphisms found in a specific patient subpopulation.

20. The method of claim 12, wherein the selected model structures represent structural variants derived from patients the receive a specific treatment regimen.

21. The method of claim 12, wherein the selected model structures represent structural variants derived from patients that exhibit a particular clinical responses to a given drug.

22. The method of claim 12, wherein the selected model structures represent structural variants derived based on the duration of a particular drug treatment.

23. The method of claim 12, wherein the structural variant models are stored in a relational database, comprising:

3-D molecular coordinates for the structural variants;

a molecular graphics interface for 3-D molecular structure visualization; and

functionality for protein sequence and structural analysis; and

database searching tools.

24. The method of claim 12, wherein the database further comprises observed clinical data associated with the genetic polymorphisms, subject medical history and subject history.

25. A computer-based method of selecting drug therapies for patients based on genetic polymorphisms, comprising:

obtaining amino acid sequences of a target protein that is the product of a gene exhibiting genetic polymorphisms, wherein the sequences represent different genetic polymorphisms;

generating 3-D protein structural variant models from the sequences;

computationally docking drug molecules with the target protein models;

energetically refining the docked complexes;

determining the binding interactions between the drug or potential new drug candidate molecules and the models; and

selecting drug therapies based on the drug or drugs that have the most favorable binding interactions with the structural variant models.

~~27~~ 26. The method of claim ~~25~~ 26, wherein the binding interactions are determined by:

calculating the free energy of binding between the protein structural variant and the docked drug molecule; and

decomposing the total free energy of binding based on the interacting residues in the protein active site.

~~28~~ 27. The method of claim 1, further after generating the 3-D structural variant models, exporting some or all of them models into a program that computationally docks the models with test compounds to assess intermolecular interactions.

~~29~~ 28. A computer-based method for predicting clinical responses in patients based on genetic polymorphisms, comprising:

obtaining one or more amino acid sequences for a target protein that is the product of a gene exhibiting genetic polymorphisms;

generating 3-D protein structural variant models from the sequences;

building a relational database of protein structural variants derived based on genetic polymorphisms and observed clinical data associated with particular polymorphisms exhibited in the patients, wherein the database comprises:

3-D molecular coordinates for the structural variant models;

a molecular graphics interface for 3-D molecular structure visualization;

computer functionality for protein sequence and structural analysis;

database searching tools; and

observed clinical data associated with the genetic polymorphisms, subject medical history and subject history associated with the genetic polymorphisms;

obtaining a target protein structural variant based on the same gene associated with a polymorphism in a patient;

generating a 3-D protein model based on the subject's gene sequence;

screening/comparing the 3-D model derived from the subject to the structures contained in the database by:

identifying structures in the database that are similar to the model derived from the subject; and

predicting a clinical outcome for the patient based on the clinical data associated with the identified structures.

328. A computer-based method for designing therapeutic agents that are active against biological targets that have become drug resistant due to genetic mutations, comprising:

obtaining a first 3-D protein structural variant model of a target protein against which a given drug has biological activity;

generating a second 3-D protein structural variant model of the target in which genetic mutations have occurred and against which the same drug is no longer biologically active;

comparing the structures of the first and second model to identify structural differences; and

performing structure-based drug design calculations in order to identify new drugs or modifications to the existing drug to bring about biological activity against the second model.

30. A computer-based method for identifying compensatory mutations in a target protein, comprising:

obtaining the amino acid sequence of a target protein containing multiple amino acid mutations that is expressed in a patient, wherein the structure of a form of the target protein that responds to a particular drug, including the active site, has been structurally characterized;

generating a 3-D structural model of the mutated protein;

comparing the structure of the mutated protein with the form of the protein that responds to the drug to identify structural differences and/or similarities arising from the mutations;

comparing the biological activities of the drug against both the mutated protein and the form of the protein that responds to the drug to determine the effects of the mutations on drug response; and

identifying the mutations in the protein that affect biological activity based on the comparisons.

31. A method for creating a 3-D structural polymorphism relational database, comprising:

obtaining one or more amino acid sequences of a target protein that is the product of a gene exhibiting a genetic polymorphism, wherein sequences represent different genetic polymorphisms;

generating 3-D protein structural variant models from the sequences;

energetically refining the models;

evaluating the quality of the models;

optionally obtaining associated clinical properties or data; and

inputting the model and any associated properties and/or data into a relational database.

32. The method of claim 31, wherein after energetically refining the models, the models are further refined.

34 ~~33~~. The method of claim ~~31~~ 32, wherein the database comprises amino sequences of two or more polymorphic variants.

35 ~~34~~. The method of claim ~~31~~ 32, wherein the database comprises amino sequences of ten or more polymorphic variants .

36 ~~35~~. The method of claim ~~31~~ 32, wherein the database comprises amino sequences of about 100 or more polymorphic variants .

37 ~~36~~. The method of claim ~~31~~ 32, wherein the database comprises amino sequences of about 1000 or more polymorphic variants .

38 ~~37~~. The method of claim ~~31~~ 32, wherein the database comprises amino sequences of more than 8000 polymorphic variants.

39 ~~38~~. A database created by the method of claim ~~31~~ 32.

40 ~~39~~. The database of claim ~~38~~ 39, comprising variant 3-dimensional structures of a selected target.

41 ~~40~~. The database of claim ~~38~~ 39 that comprises structures of proteases or polymerases.

42 ~~41~~. The database of claim ~~38~~ 39, wherein the proteases are viral proteases or polymerases.

43 ~~42~~. The database of claim ~~38~~ 39, wherein the viral proteases are human immunodeficiency virus proteases and the polymerase is a viral reverse transcriptase.

44 ~~43~~. The method of claim ~~31~~ 39 32, wherein quality is assessed by computing the normalized residue energies such that if e_{av} is $\geq$ 1.5 a model is further refined until e_{av} is $<$ 1.5; if e_{av} is $<$ 1.5 a model is deposited into the database.

45 ~~44~~. The method of claim 1, wherein the target is an enzyme.

46 ~~45~~. The method of claim ~~44~~ 45, wherein the enzyme is a protease or polymerase.

47 ~~46~~. The method of claim ~~45~~ 46, wherein the polymerase is a reverse transcriptase.

48 ~~47~~. The method of claim ~~44~~ 45, wherein the target is a protein expressed by an infectious agent.

49 ~~48~~. The method of claim ~~44~~ 45, wherein the target is enzyme expressed by a an infectious agent.

50 ~~49~~. The method of claim ~~48~~ 49, wherein the agent is a human immunodeficiency virus (HIV).

51 ~~50~~. A computer system, comprising a database containing data representative of the three dimensional structure of polymorphic variants of a drug target.

52 ~~51~~. The system of claim ~~50~~ 51, wherein the target is a cell surface receptor or an enzyme.

53 ~~52~~. The system of claim ~~50~~ 51, wherein the enzyme is a protease or a polymerase.

54 ~~53~~. A database, comprising:

sequences of nucleotides encoding a protein or portions thereof, wherein proteins comprise polymorphic variants; and the portions encode a domain of the protein that comprises a site in the protein that binds to a drug candidates; and

the coordinates of 3-dimensional (3-D) structures of the encoded proteins or portions thereof.

55 ~~54~~. The database of claim ~~53~~ 54 that is a relational database.

56 ~~55~~. The database of claim ~~53~~ 54 that comprises at least 2 polymorphic variants and the corresponding 3-D structures.

57 ~~56~~. The database of claim ~~53~~ 56 that comprises at more than 10, more than 100, more than 1000, more than 8000, or more than 10,000 polymorphic variants and the corresponding 3-D structures.

58 ~~57~~. The database of claim ~~53~~ 54, wherein the protein is a receptor or enzyme from a eukaryotic or prokaryotic organism.

59 ~~58~~. The database of claim ~~53~~ 54, wherein the organism is a pathogen or a mammal.

60 ~~59~~. The database of claim ~~53~~ 54, wherein the organism is a pathogen is a virus or bacterium and the mammal is a human.

60. The database of claim 53, wherein the protein is a protease or a reverse transcriptase.

61. A database, comprising the sequences of nucleotides set forth in SEQ ID Nos. 3-117 that encode HIV protease or the portion of HIV reverse transcriptase set forth in each SEQ ID.

62. The database of claim 53, further comprising 3-D structural coordinates for a protein or portion thereof comprising sequences of amino acids encoded by each of SEQ ID Nos. 3-117.

63. The database of claim 54, wherein the protein is HIV protease.

64. The database of claim 54, wherein the protein is HIV reverse transcriptase.

65. The method of claim 1, wherein the target protein is a eukaryotic or prokaryotic protein.

66. The method of claim 1, wherein the target protein is an animal protein, a plant protein or a protein from a pathogen.

add C1

add F22